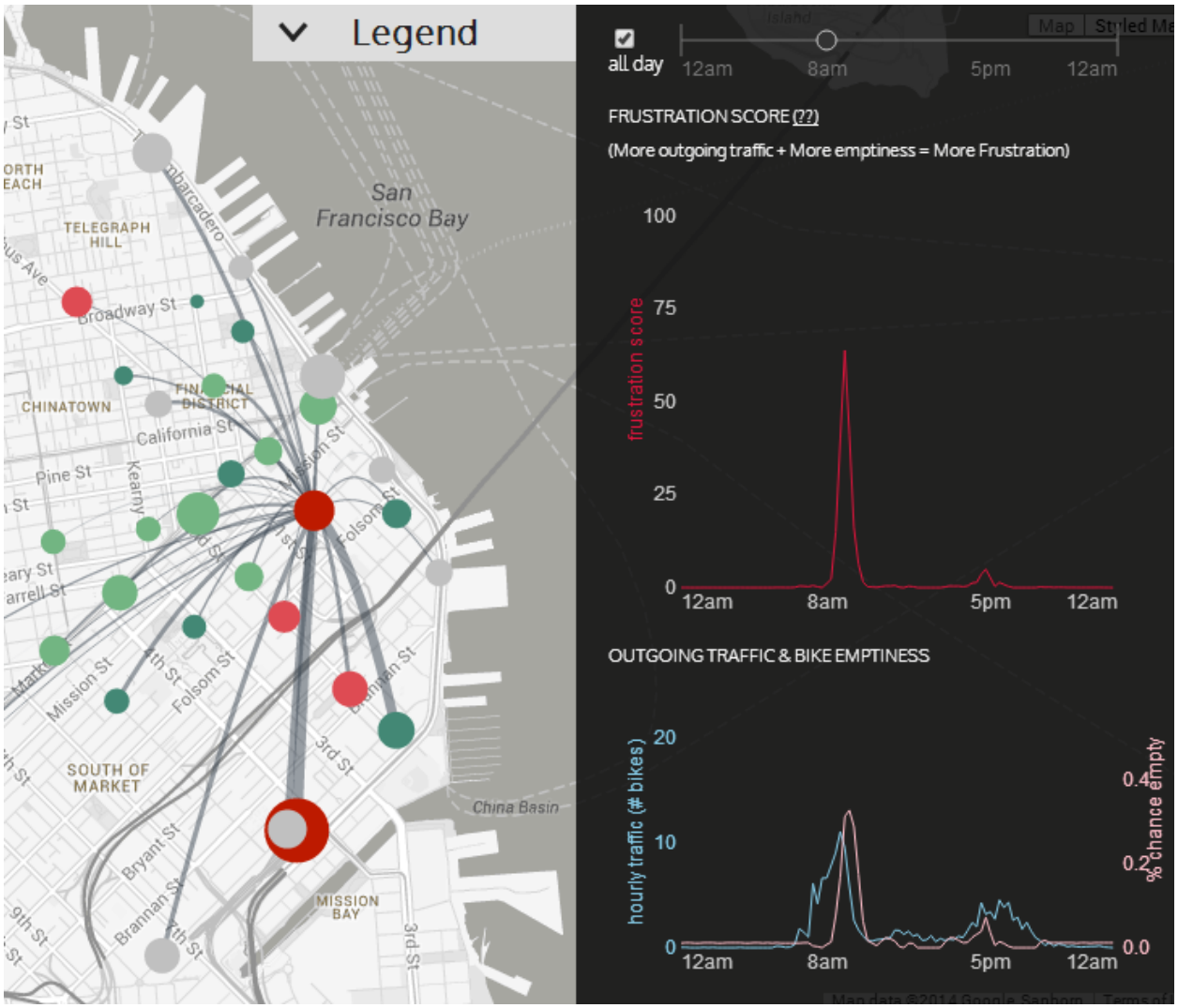
Hey, Where's My Bike?



Dan Tsai, Chan Kim, Sandra Lee
Info 247, Information Visualization
Prof. Marti Hearst

Project Goals

We are visualizing data from nearly 150,000 bike trips during the past 6 months from the Bay Area Bike Share program in the San Francisco Bay Area. Our goal is to create a visualization that allows users to easily see the volume of use across different Bike Share stations as well as the quality of service at each station. We combine these information into a visualization on a map showing how well each station meets rider demand over the course of an average day at every 15 minute interval.

Overall this project is targeted toward Bike Share users and system administrators who want to assess the use and effectiveness of the Bay Area Bike Share, particularly with regard to handling load and usage demand.

Related Work

Chicago had a similar data challenge using data from their Bike Share program, so we initially looked at the winning visualizations from that challenge to get inspiration for our project. We were also inspired by the winning submissions to the Urban Data Challenge, which visualized San Francisco Muni transit data. The idea of coming up with our own frustration metric came from one of the winning submissions that also incorporated different aspects of the data their own frustration index for riding the Muni.

Visualization Overview

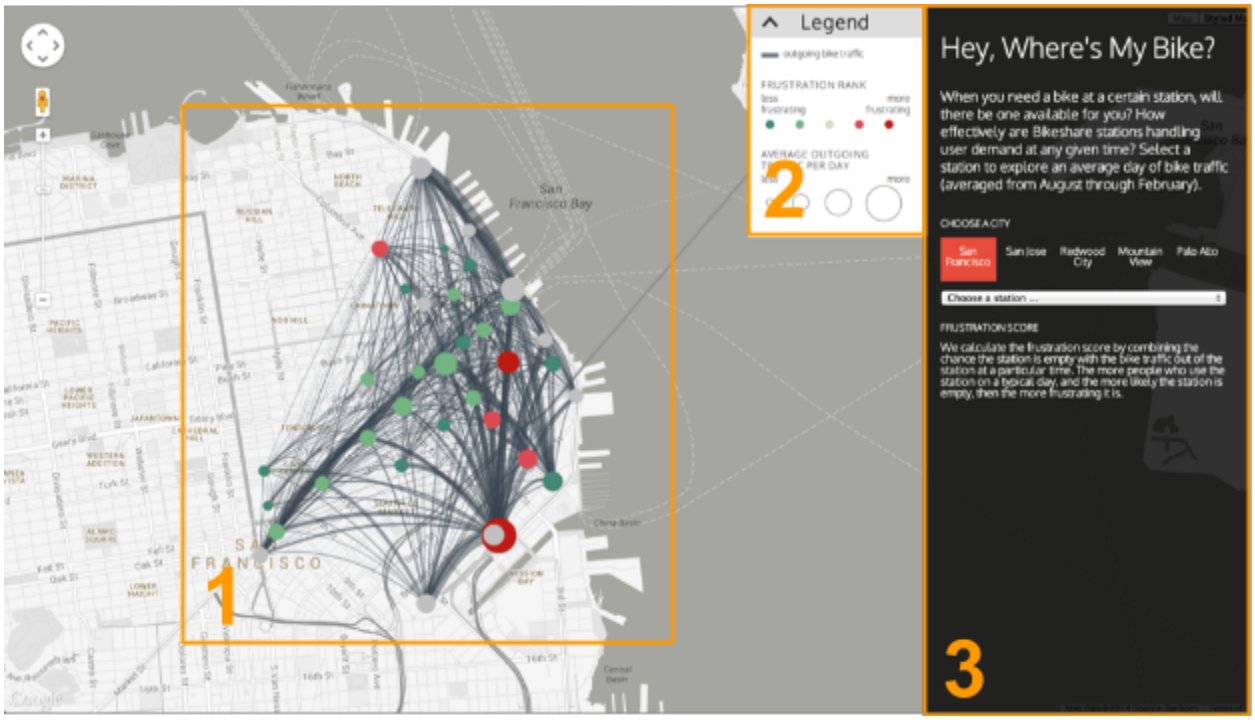


Figure 1. Landing page of ‘Hey, Where’s My Bike?’

The landing page consists of three components: 1) a map displaying the Bike Share stations and their connections 2) a legend, and 3) additional information on a side menu (Figure 1).

On the map, we draw circles for all the Bike Share stations. As shown in the legend, the size of the circles correspond to the overall volume of outgoing traffic for each station. The color of the circles represent the frustration rank we have computed for each station. Generally, the longer the station is empty, and the greater the demand, the greater the frustration is. (To read more about how the frustration score and rank are calculated, go to Exhibit B). On the map users can hover over elements (Figure 2) to see related information to the element, such as the station name, frustration rank, and average outflow from the station per day.

We also draw arcs connecting the circles to show which stations riders start at and end their trip. The widths of the arcs correspond to the volume of rides between two stations. We determined the widths by taking the minimum and maximum number of outflows among all stations and drew the arcs with the relative width.

On the side menu, we provide the app's title, an introduction, and a short description of how we calculate the frustration score. We also allow users to select one of cities which runs Bike Share program and one of stations in that city so that they can access detail information of the station which we will describe later.

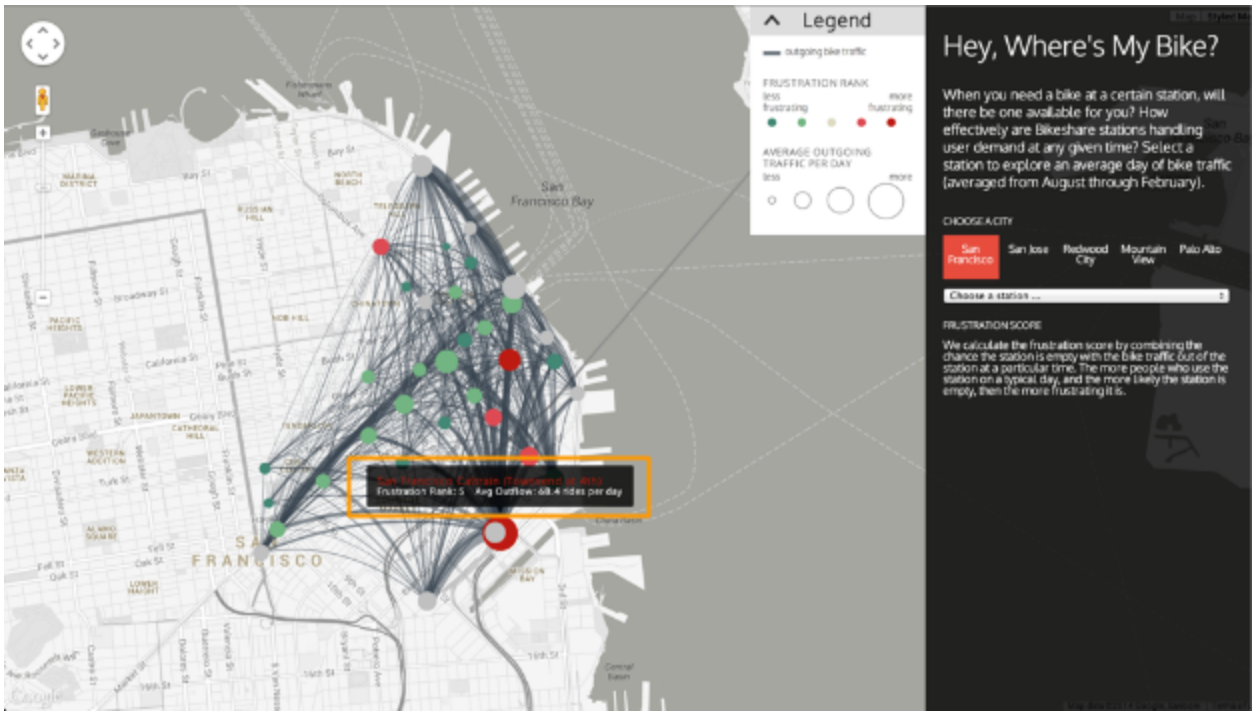


Figure 2. Tooltip for a station

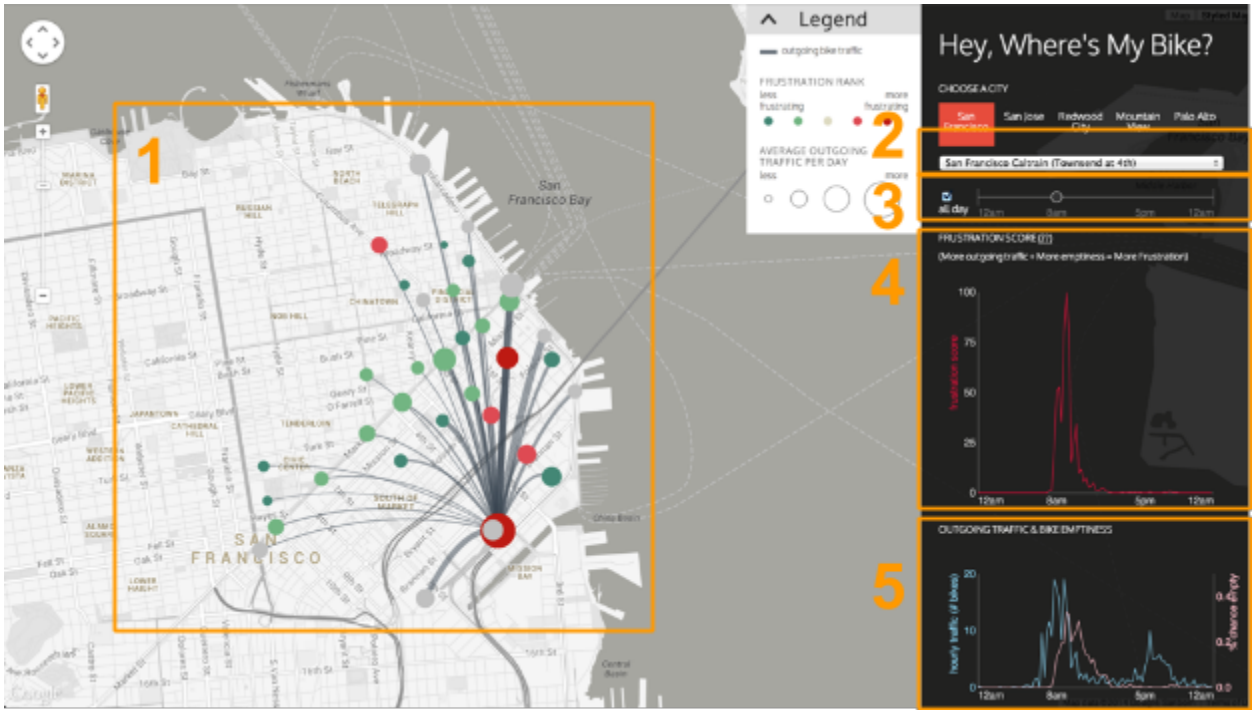


Figure 3. Detailed station information and graphs

When a user clicks on a station circle (1) or selects a station from the drop-down menu (2), the map is redrawn to show only the outgoing traffic from the selected station. The side menu also displayed more detailed information and graphs regarding that station as shown in Figure 3. One graph shows frustration scores over time (4) and the other shows average outgoing traffic over time and the percentage of time that the station has no bike (5). Users can also manipulate a time slider (3 in Figure 3) to explore the data information for a specific time (Figure 4). The color and size of the circles as well as widths of the arcs respond immediately and adjust accordingly as the user drags the time on the timeline. Users can then quickly see interesting patterns over time. For example of the Caltrain station in Figure 4, the size of the circle becomes large and the color of it turns red due to the high frustration score at 9:00am.

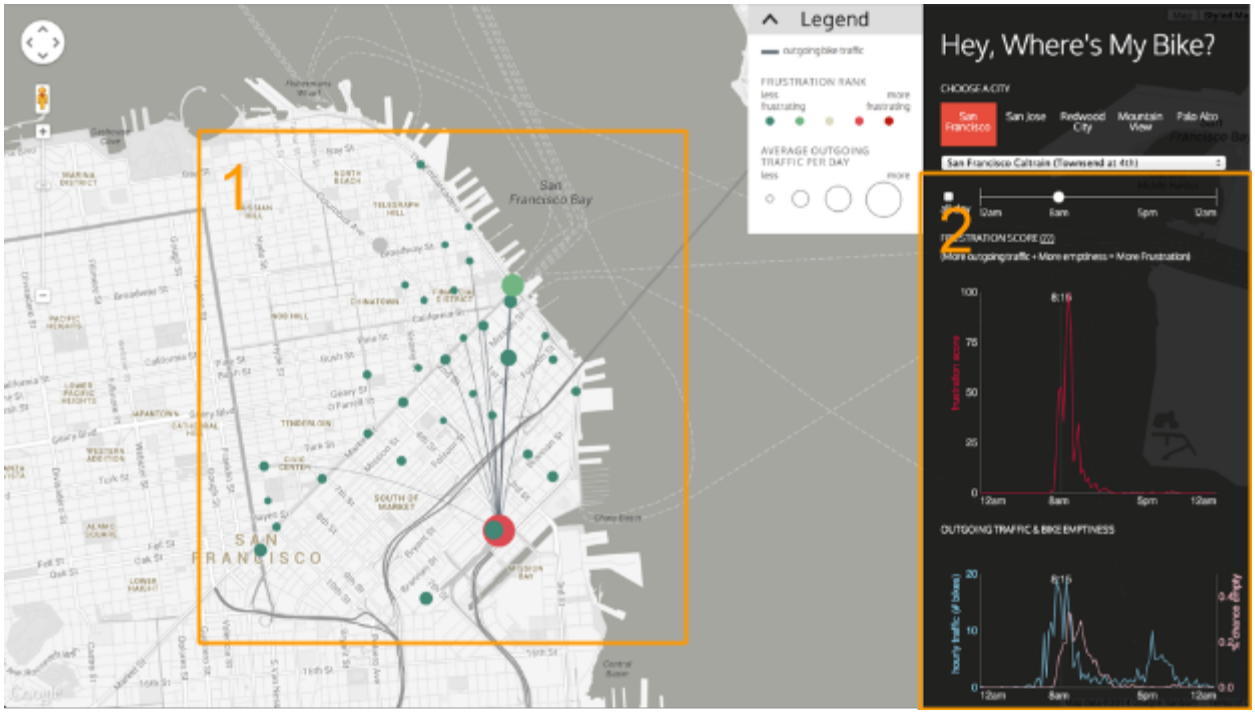


Figure 4. Detail information of a station at a specific time

After a station is selected, users can hover over the arcs to see outflow of bikes between two stations in total. (Figure 5) For those who wonder how the frustration score is calculated, we provide another tooltip explaining the process.

Additional screenshots of the visualization are attached to Appendix.

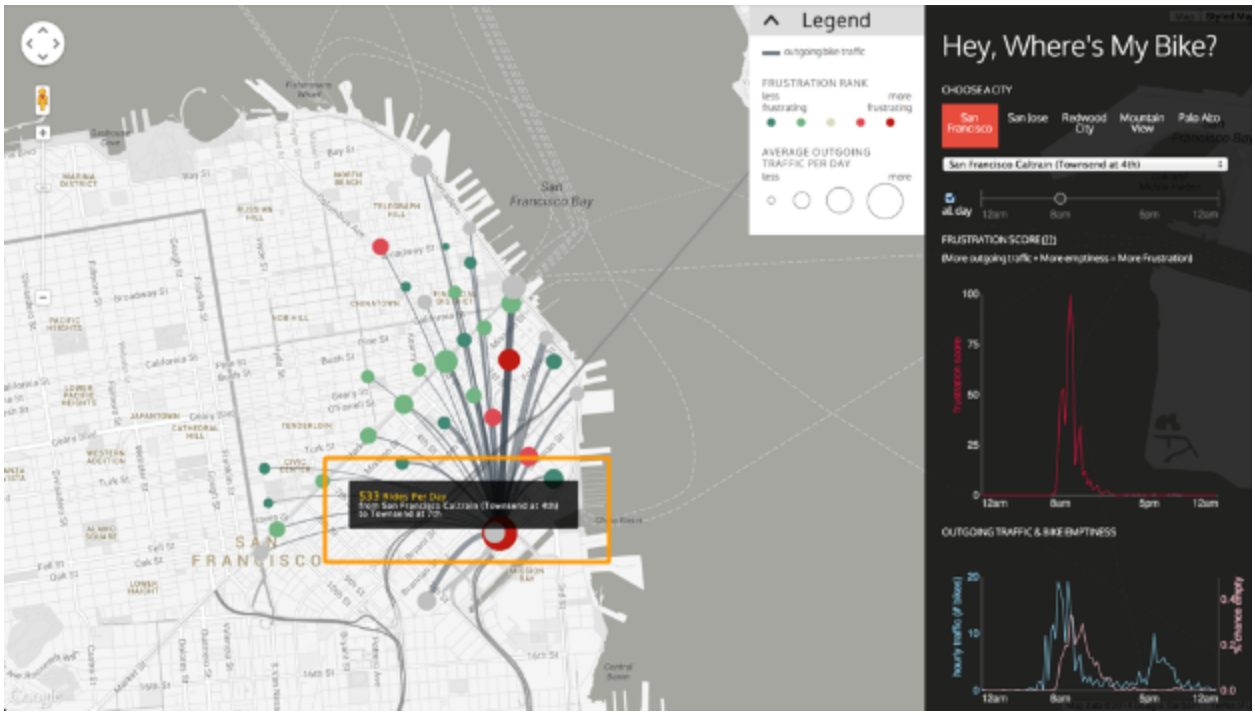


Figure 5. Tooltip for a arc

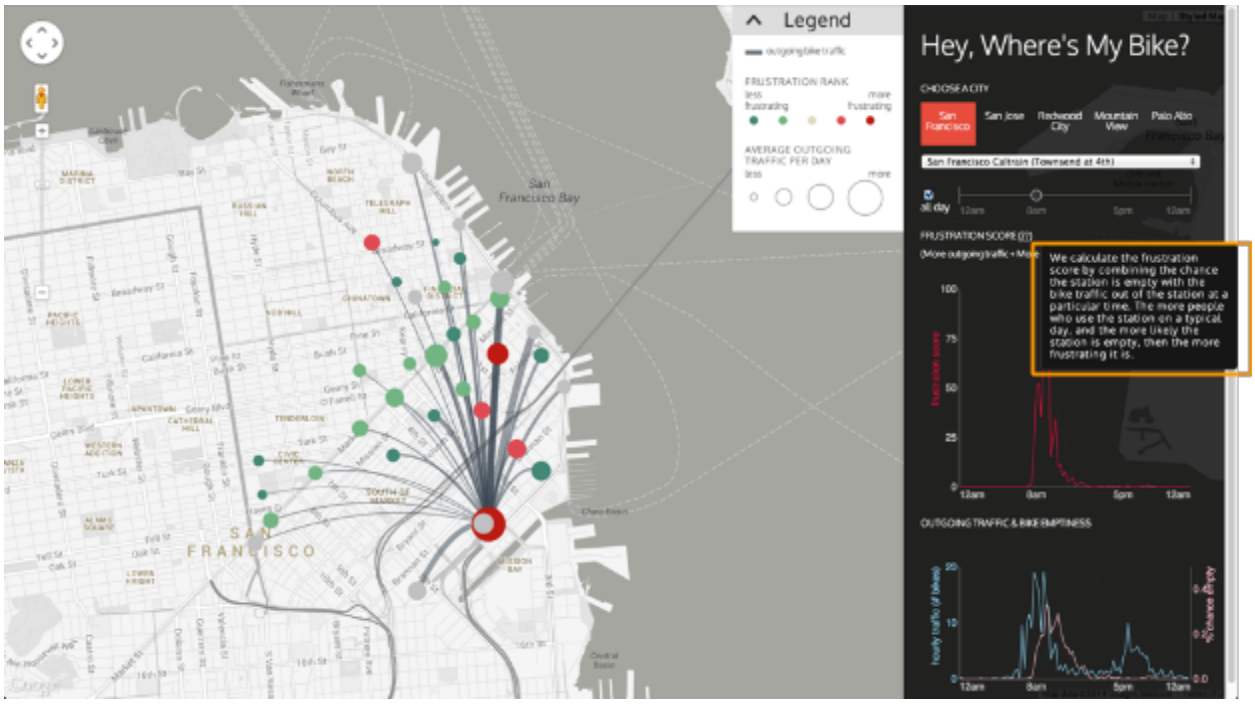


Figure 6. Tooltip for frustration score

Data

The dataset came from the Bay Area Bike Share Data Challenge. This includes latitude/longitude data for all the bike stations,all trip data within a 6 month period (time, starting station, ending station), and minute-by-minute data for all the docks and bikes available at each station.

Tools

The front end of our visualization was coded using. HTML, CSS, and JavaScript. We made extensive use of the d3.js and jQuery frameworks for interacting with the DOM and drawing the graphs and on the map. We also made use of the Google Maps API.

Design Process

We started by doing some exploratory analysis of the data using Microsoft Excel and Tableau. We then discussed various ideas before settling on the current "frustration score" idea.

Because the deadline for the Bike Share Data Challenge was in late April, we had to quickly build an initial version of our application to turn in. This version did not include the timeline function, pretty arcs, or tooltips when hovering over a station. Due to the crunched timeline, we wished we had more time to explore

alternative ideas and questions, as well as explore the data more extensively.

We started by first processing and transforming the data into JSON files using Python scripts. For our initial version we used Google Map's API to draw the circles and lines. We found that this limited us in the types of elements we could draw, such as arcs. We found it challenging to use d3 to draw SVG elements on top of the Google Map and have the coordinates still line up on the map after zooming and panning. After some trial and error, we finally succeeded in mapping lat/long coordinates to pixels on the map that were redrawn correctly whenever the map was panned or zoomed. We then redrew the lines as nicer looking arcs connecting the stations. We then added the timeline feature that redrew the map as users selected times. Finally we added tooltips upon hovering over the stations of arcs.

Results and Usability Testing

Since our visualization is targeted to both system administrators and bike share users, we conducted usability tests a representative from each of our target audience groups. The first was a former Bike Share user and software developer who critiques UI's for a living. Overall, he liked the clear explanation and purpose of the site and found the frustration score "nicely branded" and easy to understand. However, he didn't find the actual frustration score to be that useful for him as a bike user. He mentioned that an actual bike user would care less about user demand and just focus on the percentage chance that a station is empty. But it is very unlikely that bikers would change their commute time or route unless there was a very high chance of emptiness. He also gave us valuable information about the Bike Share that we did not know about, such as mentioning that there are info screens at each station showing stations nearby that have available bikes.

The second tester is a PhD student at the Department of City and Regional Planning who studies pedestrian and bicycle behavior. He gave us input from the perspective of a system administrator of the Bike Share program. He thought our frustration score was useful but mentioned that the term "frustration" would be more appealing to bike share users and not to system administrators who would probably name that term "supply/demand score" instead. He was also more interested in slicing down the data into as many dimensions as possible to explore all the nuances in the ridership, such as differences in weekend/weekday usage or impact of holidays on ridership.

Overall, because we targeted two very different audiences for our visualization, the feedback about the usefulness of our visualization was mixed. Bike share users are not all that interested in supply and demand as long as they can get a bike, while system administrators would want to explore as much of the data as possible. In the end, we learned the important lesson of doing user research much earlier in our design process. But after incorporating all the feedback and reassessing our final visualization, we felt that our project did succeed in answering this specific question: "Given my commute, should I consider a Bike Share membership?"

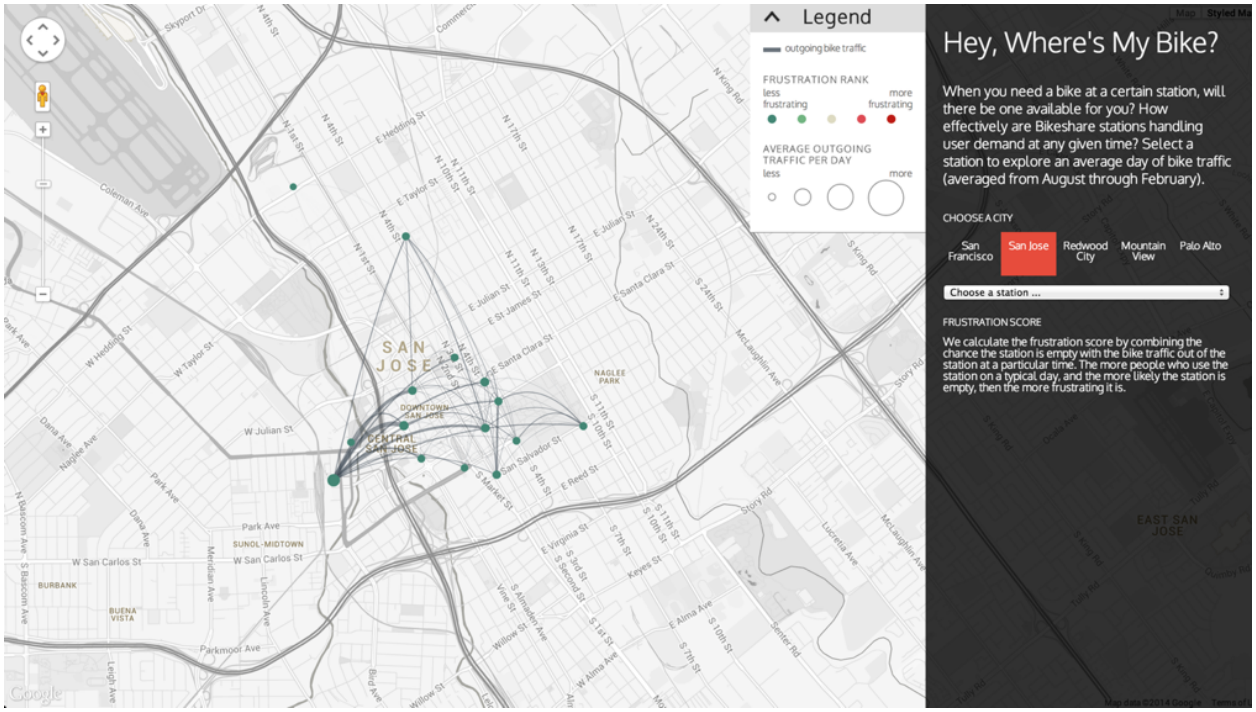
Demo and Source

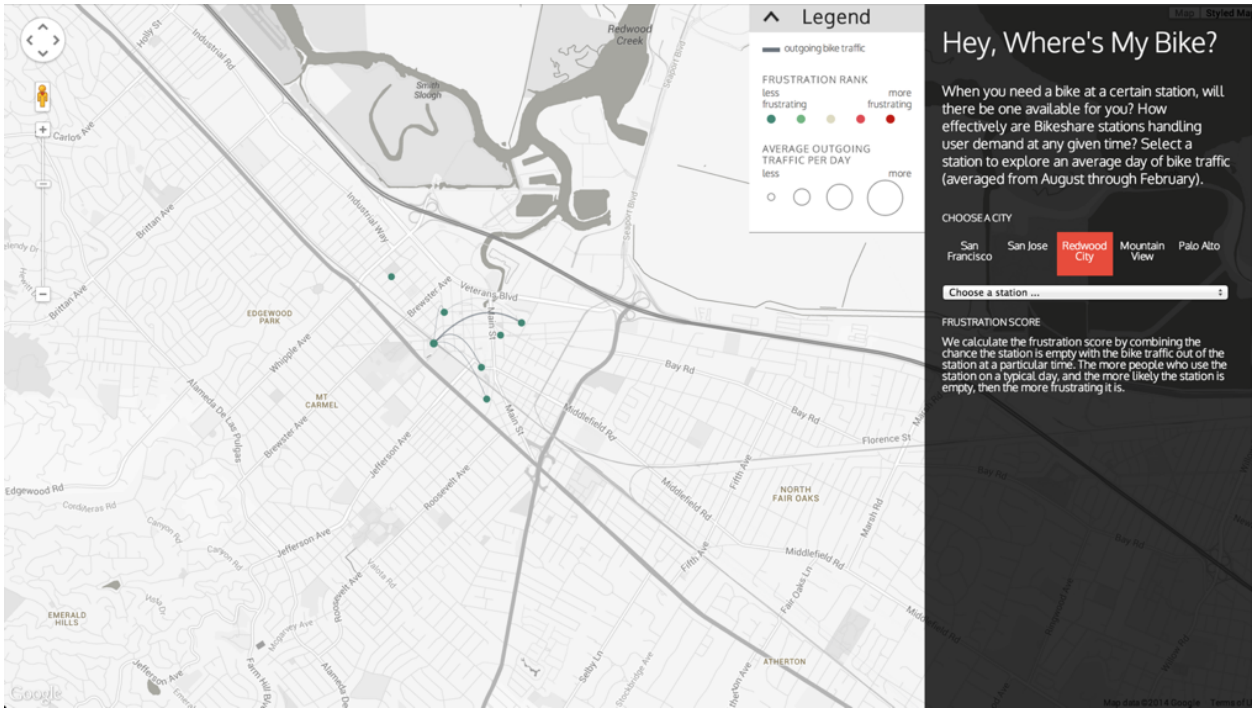
A demo of the visualization is available at http://dantsai.com/wheresmybike
Link to github repository: https://github.com/96chan/bikeshare

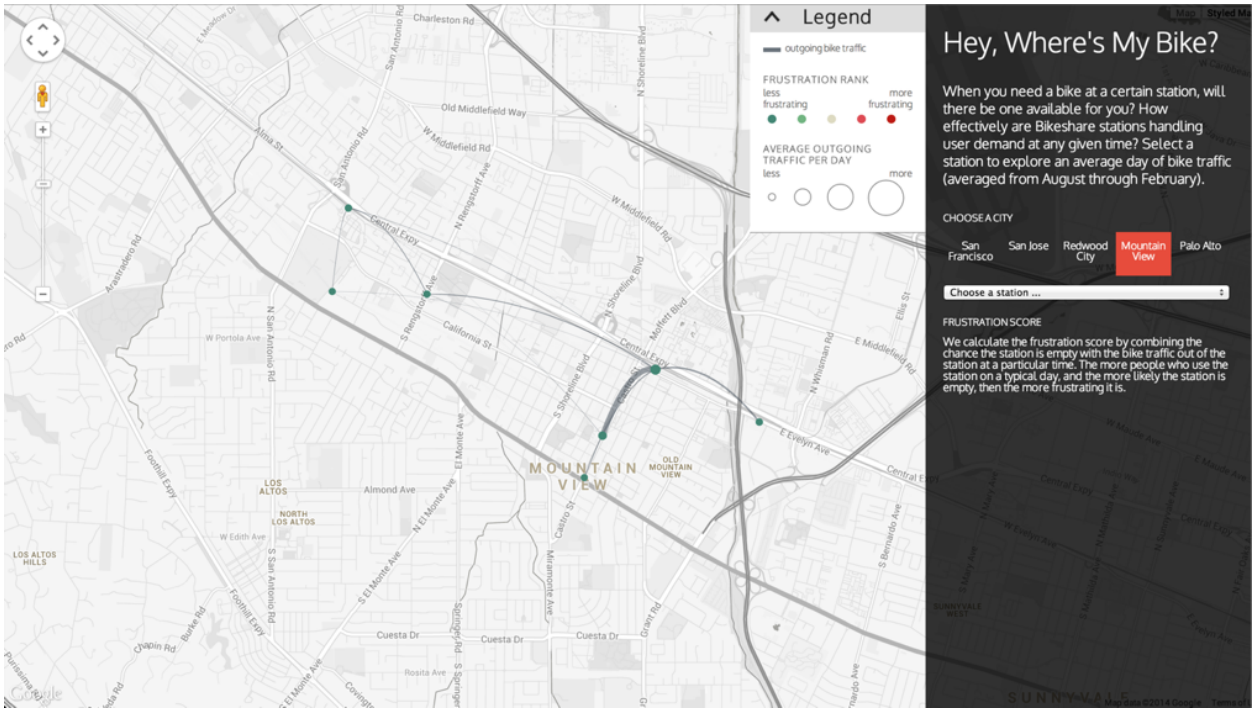
Division of Tasks

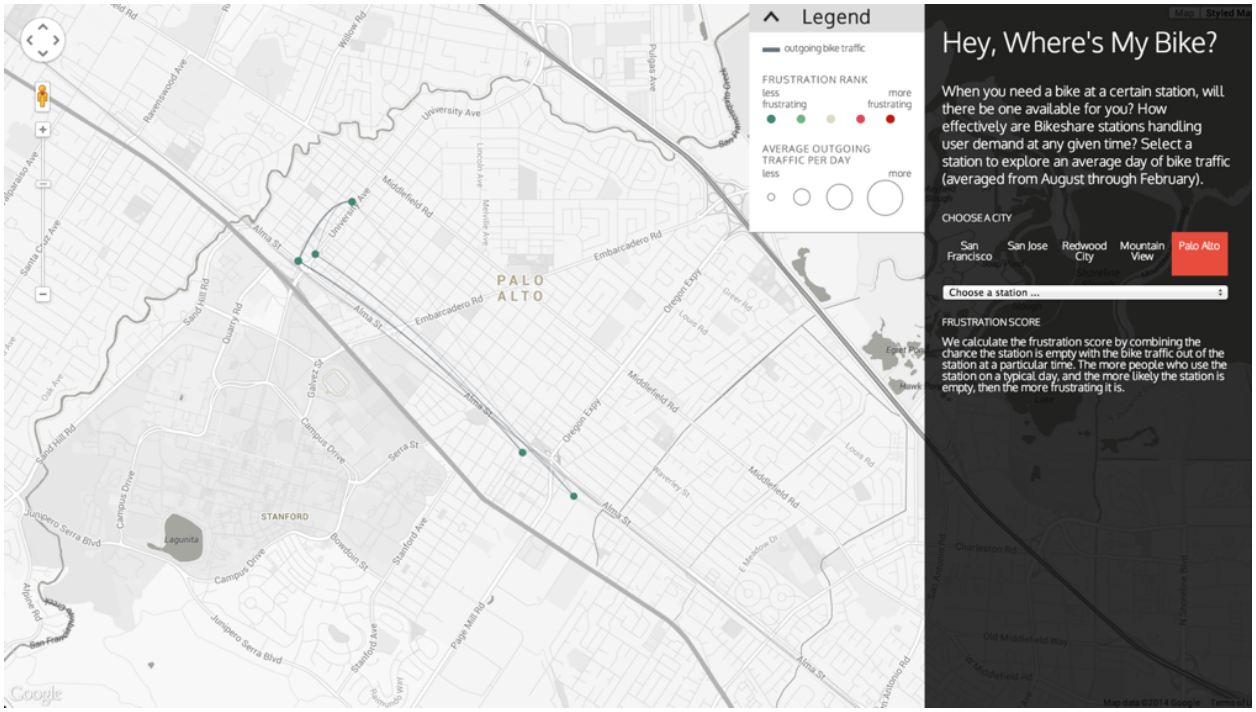
Task	Chan	Dan	Sandra
Initial Brainstorming (including Exploratory Data Analysis)	33%	33%	33%
Coming up with calculation of Frustration Score	20%	40%	40%
Data Processing and Preparation	0%	0%	100%
Website layout and styling	50%	40%	10%
Drawing lines and circles (using Google Maps API) for the initial data challenge	100%	0%	0%
Drawing arcs and circles (using Google Maps API and d3)	40%	50%	10%
Drawing graphs on the side bar	0%	95%	5%
Timeline selection, and linking with map	0%	100%	0%
User Testing	0%	0%	100%
In Class Presentation	20%	40%	40%
Final Report	33%	33%	33%

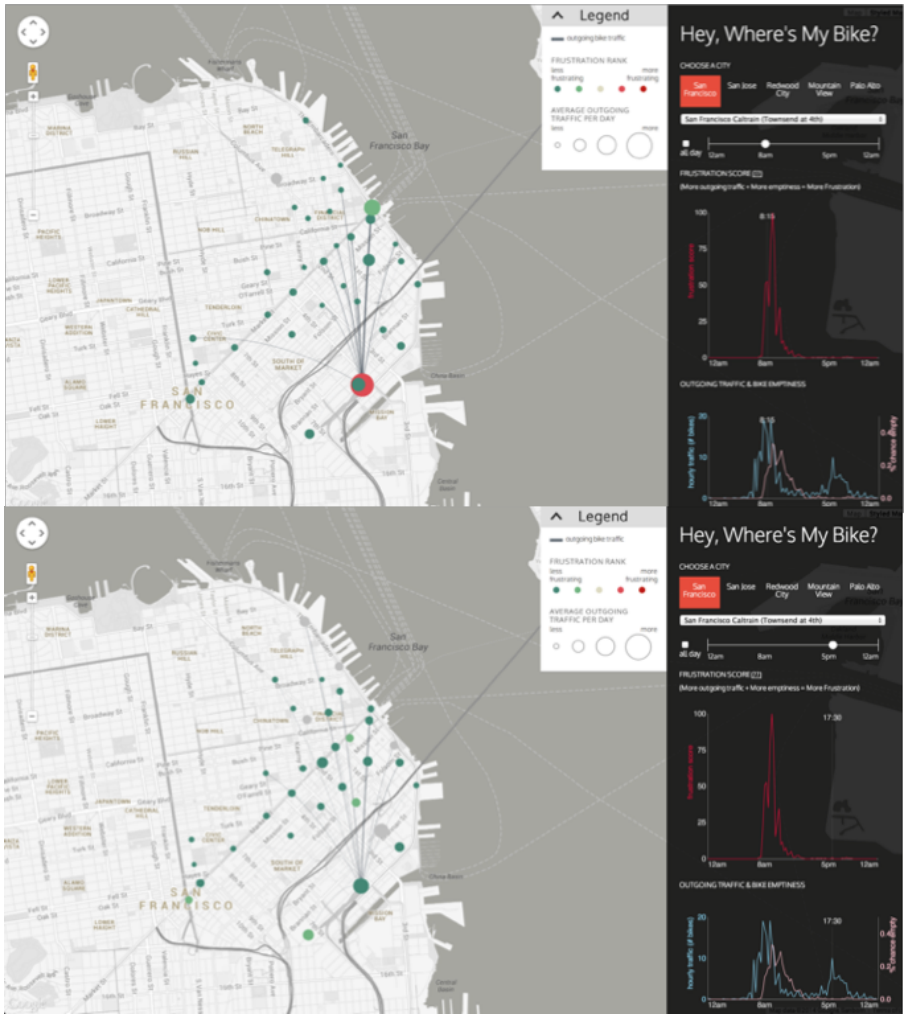
Appendix A: Screenshots


Legend
outgoing bike traffic
FRUSTRATION RANK
less frustrating
more frustrating
AVERAGE OUTGOING TRAFFIC PER DAY
less
more
Hey, Where's My Bike?
When you need a bike at a certain station, will there be one available for you? How effectively are Bikeshare stations handling user demand at any given time? Select a station to explore an average day of bike traffic (averaged from August through February).
CHOOSE A CITY
San Francisco
San Jose
Redwood City
Mountain View
Palo Alto
Choose a station ...
FRUSTRATION SCORE
We calculate the frustration score by combining the chance the station is empty with the bike traffic out of the station at a particular time. The more people who use the station on a typical day, and the more likely the station is empty, then the more frustrating it is.
SAN JOSE
DOWNTOWN SAN JOSE
CENTRAL SAN JOSE
NAGLEE PARK
SUNOL-MIDTOWN
BURBANK
BUENA VISTA
EAST SAN JOSE
Google


Legend
outgoing bike traffic
FRUSTRATION RANK
less frustrating
more frustrating
AVERAGE OUTGOING TRAFFIC PER DAY
less
more
Hey, Where's My Bike?
When you need a bike at a certain station, will there be one available for you? How effectively are Bikeshare stations handling user demand at any given time? Select a station to explore an average day of bike traffic (averaged from August through February).
CHOOSE A CITY
San Francisco
San Jose
Redwood City
Mountain View
Palo Alto
Choose a station ...
FRUSTRATION SCORE
We calculate the frustration score by combining the chance the station is empty with the bike traffic out of the station at a particular time. The more people who use the station on a typical day, and the more likely the station is empty, then the more frustrating it is.
Redwood Creek
EDGEWOOD PARK
MT CARMEL
NORTH FAIR OAKS
EMERALD HILLS
ATHERTON
Google


Legend
outgoing bike traffic
FRUSTRATION RANK
less frustrating
more frustrating
AVERAGE OUTGOING TRAFFIC PER DAY
less
more
Hey, Where's My Bike?
When you need a bike at a certain station, will there be one available for you? How effectively are Bikeshare stations handling user demand at any given time? Select a station to explore an average day of bike traffic (averaged from August through February).
CHOOSE A CITY
San Francisco
San Jose
Redwood City
Mountain View
Palo Alto
Choose a station ...
FRUSTRATION SCORE
We calculate the frustration score by combining the chance the station is empty with the bike traffic out of the station at a particular time. The more people who use the station on a typical day, and the more likely the station is empty, then the more frustrating it is.
MOUNTAIN VIEW
LOS ALTOS
NORTH LOS ALTOS
LOS ALTOS HILLS
Google


Legend
outgoing bike traffic
FRUSTRATION RANK
less frustrating
more frustrating
AVERAGE OUTGOING TRAFFIC PER DAY
less
more
Hey, Where's My Bike?
When you need a bike at a certain station, will there be one available for you? How effectively are Bikeshare stations handling user demand at any given time? Select a station to explore an average day of bike traffic (averaged from August through February).
CHOOSE A CITY
San Francisco
San Jose
Redwood City
Mountain View
Palo Alto
Choose a station ...
FRUSTRATION SCORE
We calculate the frustration score by combining the chance the station is empty with the bike traffic out of the station at a particular time. The more people who use the station on a typical day, and the more likely the station is empty, then the more frustrating it is.
PALO ALTO
STANFORD
Google



Appendix B: Frustration Score Methodology

The frustration score meant to capture the demand and lack of supply of bicycles. Demand is reflected by the average outgoing rides per day at a station. Lack of supply is reflected by the percentage of time that the station is empty. To calculate the score, we multiply the average outgoing traffic for each station by the percentage of time that the station is empty for each 15 minute interval. We first calculate an individual frustration score for each station at each 15 minute interval. We then normalize the scores from 0-100 scale according to the highest and lowest scores. The normalized scores are shown on the frustration graph. To display the score by color on the map, we ranked all the scores into 5 categories using the Jenks natural breaks classification algorithm and color coded them accordingly.

To calculate the overall frustration scores for a station (as shown on the landing page), we take the average of all the individual frustration scores that we had previously computed for every 15 minute interval of that station. We then normalize these scores and again apply the Jenks natural breaks classification to rank these scores into 5 categories.